Simulations for the next Generation of Large-Scale Structure Surveys

Risa Wechsler
(Stanford/SLAC/KIPAC)

Cosmic Visions @ SLAC
November 13, 2015

simulations are at the heart of modern cosmological analysis

- Predictions
  - Accurate predictions of the mean signal as a function of cosmological model
  - Accurate predictions of covariances between observables
- Understanding systematics
  - Developing and testing analysis pipelines
  - Characterizing and learning to marginalize over data systematics
  - Characterizing and learning to marginalize over astrophysical systematics

next generation surveys

- currently happening: DES -- first experiment that will measure all of the “key” DE probes in one survey.
  - still on a steep learning curve about how to do these analyses most effectively, still learning which systematics will dominate
- DESI: 35 million galaxies and quasars to z ~ 3.5
  - BAO + RSD allow same sample to probe expansion history and growth of structure
  - significant additional power / discovery space in joint probes & small scales, including densely-sampled BGS at low z
- LSST: 10+ billion galaxies
  - many probes no longer statistics limited, need exquisite control of both data and theory systematics
  - what is the range of scales we will use for cosmology?
- generally:
  - large volumes, large dynamic range of galaxy properties, scales
  - precise measurements -- better than 1% measurements for many statistics
  - most potential comes from complementarity of several probes

$\longrightarrow$ places strong requirements on modeling

- General challenges:
  - defining and simulating the full range of cosmological models
  - simulating the full range of scales over the volumes that will be measured
  - want predictions for observables, not just clean quantities of the matter distribution from gravity-only simulations
  - data management -- can be more challenging than for the surveys themselves, because we need multiple simulated sky surveys

cosmological models

- simulating LCDM is straightforward, but still cannot get full dynamic range desired in one simulation
- dark energy models (e.g. w(z)CDM)
  - easy to implement, but so far neglects perturbations
- modified gravity
  - what is the model space we are probing?
  - very expensive, distinguishing power is on non-linear scales
- dark matter & neutrinos
  - simulations with neutrinos still not doing the full calculation -- are they accurate enough?
  - some progress with self-interacting DM, but expensive and still some modeling uncertainties

accurate predictions

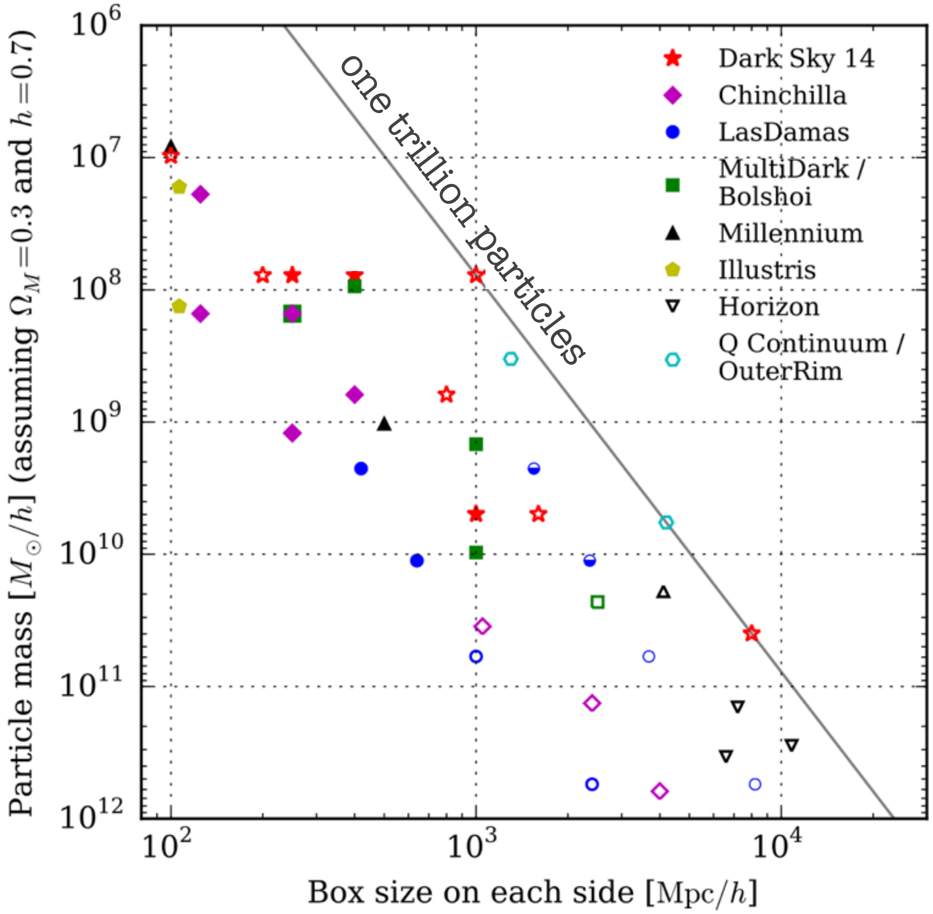
- predictions for basic statistics
  - matter power spectrum
  - mass and bias functions for dark matter halos
  - can build fast emulators from a set of several tens of simulations (depending on parameter space, accuracy needs)
  - current accuracy of models/emulators: 5-10% (1% in some regimes), need to get to ~1% for wide range of models and scales
- need to do this for range of observed statistics
  - galaxy statistics (clustering, RSD, etc)
    - need detailed understanding of how galaxies trace the dark matter, or ability to marginalize over the parameters that specify it
  - more complex statistics: e.g. cross-correlation of galaxies with CMB, galaxies with clusters, galaxies with troughs, lensing x CMB lensing etc...
- need to do this over the full range of cosmological parameters / models that we want to test

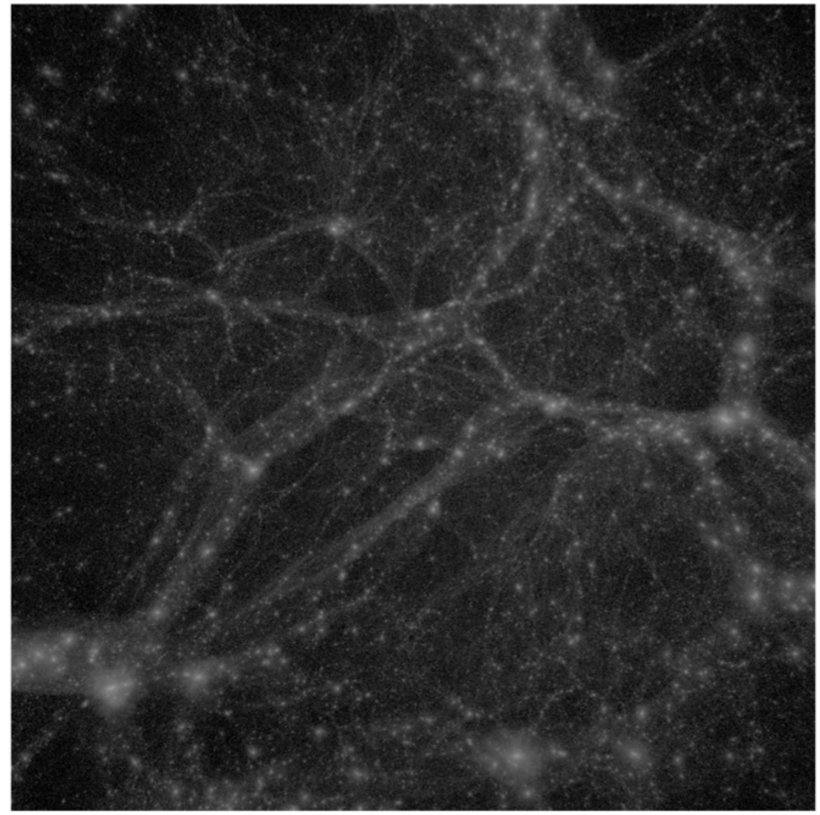
covariances

- Rule of thumb is
  - if you estimate the covariance between N_D data points with N_S simulations, variance in parameters increases by $1+N_D/N_S$
  - ~ 5x sims per data point so that errors increase less than 10%
  - want to include realistic survey details in these calculations, so ideally want to make a large number of full survey realizations
- Brute force numbers are very high!
  - need to work on data compression
  - potential for new techniques to reduce the needed number of simulations -- use smaller number to tune theoretical calculations instead of brute force calculations with the full number of simulations

Fundamental limit of how much cosmological information we can extract from these surveys will be our ability to model and understand systematics over the full survey area and the accuracy with which we can do this to small scales

current N-body simulations





trade-offs between volume, resolution, cosmological model space, included physics.

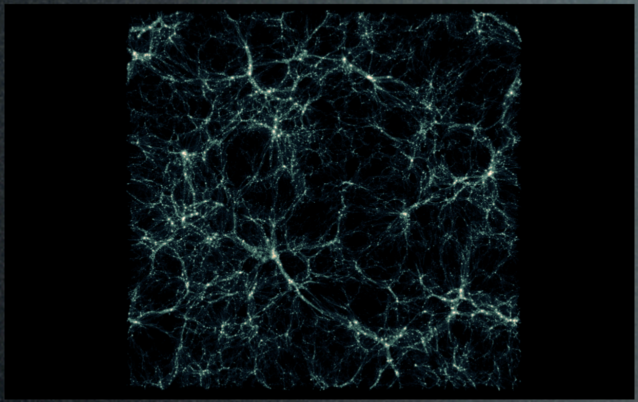
simulated sky surveys

- Test full analysis pipelines on simulations that are as realistic as possible to understand and calibrate systematics
- Model realistic galaxy populations as well as survey details
- Ideally want to model all of the major cosmological probes for a survey in one simulation
    - observed properties of galaxies
    - large-scale structure of galaxies
    - realistic impact of lensing shear on galaxies
    - galaxies in the sky vs. galaxies on our CCDs and in our catalogs -- as many relevant observational systematics as possible
- Want to produce many full area and depth sky surveys
    - many cosmological models
    - a variety of galaxy models for a given cosmology
    - multiple skies for covariance

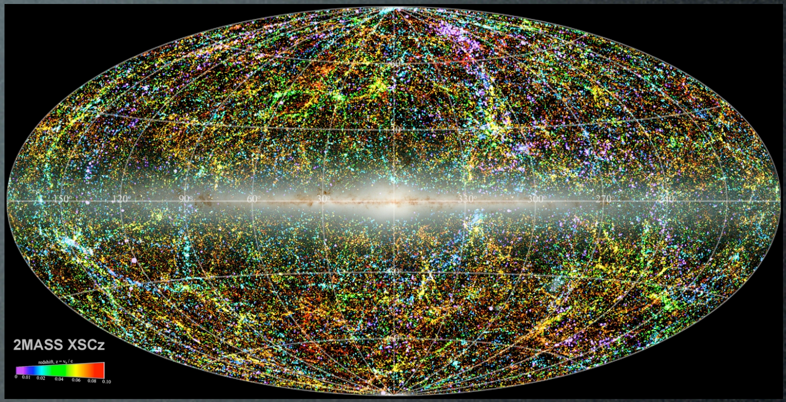
systematics in these surveys depend on galaxies

- We are moving from a statistics-limited regime to a systematics-limited regime --> need accuracy, not just precision!
- Systematics in making the map from an imaging survey
  - photometric redshifts, calibration, dust, star-galaxy separation, deblending, etc etc.
- Systematics in making robust predictions for a given model
  - non-linear structure formation
  - how galaxies trace the matter distribution
  - impact of galaxy formation physics on the power spectrum
  - intrinsic alignments

modeling galaxies



→



- ideal: predict the galaxy population for a given cosmological model e.g. P(k | L, SFR, color, etc) from first principles

  not actually (yet) possible! accuracy with which we can do this is improving but unlikely to keep up with the improving accuracy of our measurements

- practical: describe the galaxy population for a given cosmological model with a flexible parameterization; marginalize over this parameterization for the possible galaxy population when constraining cosmology

- + lots of possibilities in between...

some challenges

- large dynamic range -- difficult to get volume and accuracy at the same time
- most models are computationally expensive, so it's hard to explore parameter space
- models are not general enough -- even when exploring parameter space, they don't match the data to the precision with which it is measured
- models do not capture all of the relevant physical processes, and thus may be missing some of the relevant correlations

approaches to and considerations in galaxy modeling

- hydro sims
- semi-analytic models
- empirical models
  - lots of flavors
  - HOD/CLF/SHAM/CAM/ADDGALS
- key distinctions:
  - are galaxy properties determined by direct simulation, by calculation from simplified physical recipes, or from empirical techniques?
  - are all galaxies in resolved host halos?
  - are all galaxies in resolved subhalos?
  - do galaxy properties depend on merger history?

- how much volume?
  - surveys now probing several tests of Gpc3, ideally would like to simulate survey volumes many times, many cosmologies
- how faint?
  - typically cosmology probes use >0.1L* galaxies, but fainter galaxies matter e.g. for lensing source population
- what galaxy properties are important?
- what correlations are important?
  - e.g. correlations between galaxy properties and assembly history and large-scale structure

simulations requirements are highly dependent on the answers to these questions

modeling all the galaxies

- modeling advances
  - significant progress in simulating realistic galaxies with hydro, but still quite basic uncertainties -- can we fully bracket the range of things that baryons can do?
  - significant progress in empirical models -- but need to make sure they are flexible enough and yet still predictive.
- how far can we push it?
  - want to fully use the predictive power of simulations where they are robust
  - need models that are flexible enough to encompass the uncertainties, and need to be sure that they include the relevant correlations
  - down to what scale and accuracy level can we make these predictions?
- tiered strategy
  - use empirical models based on resolved halos and subhalos, that are tested to encompass range of possibilities that occur in the best physical models
  - extend these models to lower resolution simulations to get full volume of surveys
  - constrain these models with data jointly with cosmological parameters

simulating full surveys

simulation

halo finding

merger trees

assign galaxies

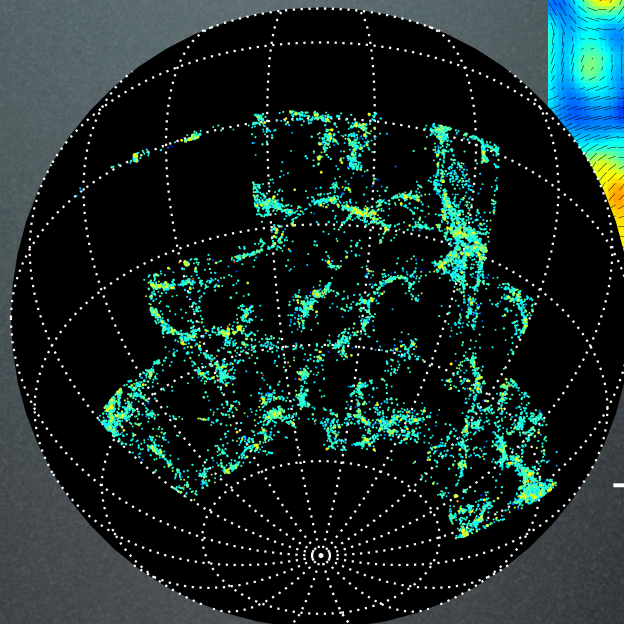
calculate shear

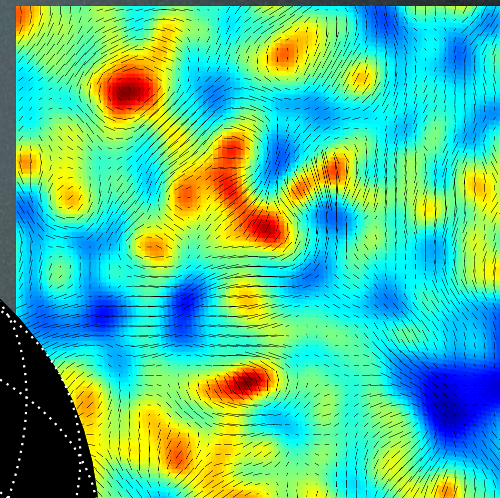
lens galaxies

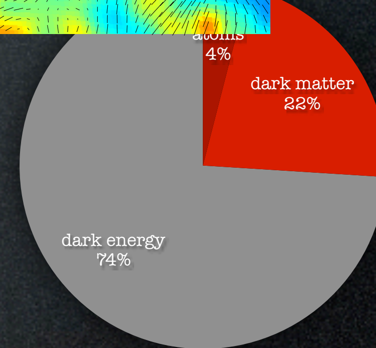
photometric errors

photometric redshifts

create sky images







Produces simulated catalogs of ~1 billion galaxies (i ~ 26) over 1/4 of the sky, on a lightcone out to z=2

Includes galaxy photometry in many bands, galaxy magnitudes and shapes fully lensed along the lightcone

Extensively tested with SDSS, DES, etc.

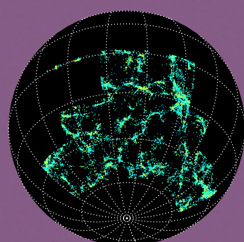
Wechsler, Becker, deRose
with DES simulation working group

DES Mock Pipeline / R. Wechsler

example simulation effort for DES

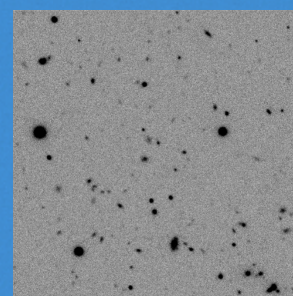
Cosmological simulations + galaxy populations + lensing

(Becker, Busha, DeRose, RW)



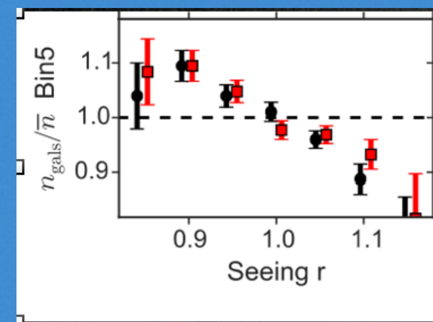
UFIG imaging simulations

(Chang et al 2015)



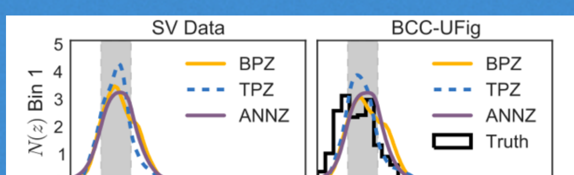
Systematics maps

(Leistedt et al 2015)



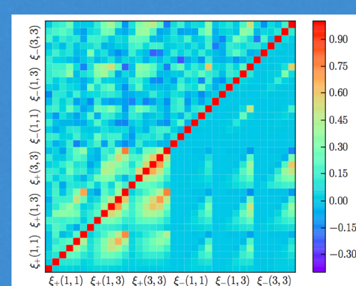
Photometric Redshifts for Lensing

(Bonnett et al 2015)



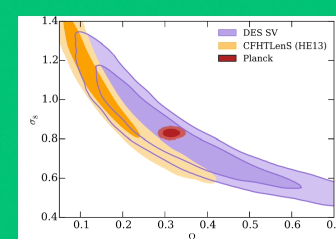
Shear 2pt Covariances

(Becker et al 2015)



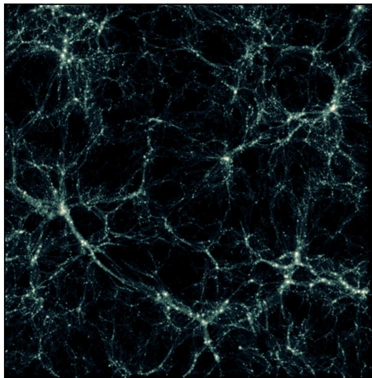
Cosmology Constraints from Shear 2-pt

(DES collaboration 2015)

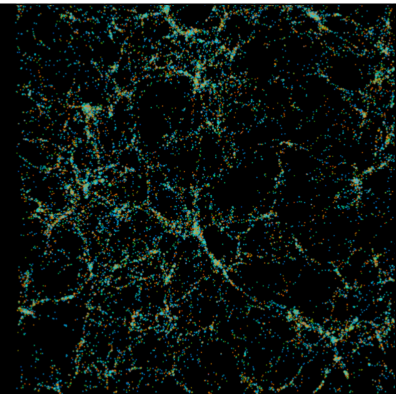


simulations are at the heart of modern cosmological analysis

- Accurate predictions of the mean signal, as a function of all cosmological models to be tested
    - e.g., matter power spectrum
    - galaxy statistics (need detailed understanding of how galaxies trace the dark matter)
    - more complex statistics: e.g. cross-correlation of galaxies with CMB, galaxies with clusters, galaxies with troughs, lensing x CMB lensing etc...
- Understanding covariances between observables
- Astrophysical systematics
    - e.g. impact of galaxy formation on matter power spectrum, intrinsic alignments, etc.
- Data systematics
    - e.g. photometric redshifts, calibration, dust, star-galaxy separation, deblending, etc etc -- simulations can provide realistic correlations with underlying structure
- Testing analysis pipelines
    - especially important for complex cosmological analyses that involve multiple probes and complex systematics



Summary



- Extracting accurate parameters of interest out of cosmological surveys increasingly requires simulations in many aspects of the analysis, including for the basic predictions and for understanding systematics.
- The forefront of this work will be bringing the predictions all the way to the observables -- how accurately can we do this, and over what range of scales?
- Diverse simulation effort needed: gravity only, detailed galaxy modeling, hydrodynamics.
- Major effort analogous to high-energy physics Monte-Carlo simulation program. Effective investment in simulation effort (both computing & people) can maximize investment in and significantly improve the science reach of currently planned cosmological surveys.